Received: 12 May 2019 | Revised: 21 July 2019 | Accepted: 5 August 2019 | Published on: 15 October 2019
DOI: 10.1002/asna.201913665

ORIGINAL ARTICLE

Astronomische Nachrichten

Methods for identifying high-redshift galaxy cluster candidates

Sandor Pinter[1,2] | **Lajos G. Balázs**[2,3] | **Zsolt Bagoly**[4] | **Istvan Horvath**[1] | **István I. Rácz**[1,4] | **Viktor L. Tóth**[2]

[1]Department of Natural Science, National University of Public Service, Budapest, Hungary

[2]Department of Astronomy, Eötvös University, Budapest, Hungary

[3]MTA CSFK Konkoly Observatory, Budapest, Hungary

[4]Physics of Complex Systems Department, Eötvös University, Budapest, Hungary

Correspondence
Sandor Pinter, Department of Natural Science, National University of Public Service, Budapest, Hungary.
Email: s.pinter@astro.elte.hu, sandor.pinter@uni-nke.hu

Present address
Sandor Pinter, National University of Public Service, Budapest, Hungária krt. 9, Budapest 1101, Hungary.

Funding information
OTKA grant, K131653; TKP Program; Tisza István Program

Abstract
Recent theories linked long gamma ray bursts (GRBs) to galaxies with rapid star formation or starburst; thus, we expect that long GRBs (LGRBs) are more frequent in midcluster galaxies where mergers and tidal interactions between gas-rich galaxies are more likely to occur. Yet there is no galaxy cluster known to be associated with LGRBs. We demonstrate that, based on deep, single-band Subaru Hyper Suprime-Cam observations, we may provide constraints on photometric redshifts of groups of galaxies. We compare three methods: cosmological approach, pseudoinverse matrix, and random forests to estimate galaxy and quasar redshifts. Comparing our results to spectroscopic redshifts of Sloan Digital Sky Survey's-detected extragalactic sources, random forests may provide the highest accuracy with as low as 17 percentage error. This is a powerful method to find clusters to place GRB host galaxies in their local environment.

KEYWORD
galaxies: distances, clusters – methods: statistical – techniques: image processing, photometric

1 | INTRODUCTION

To better understand the phenomena of gamma ray bursts (GRBs) and the physical properties of the progenitors in high-redshift GRB Host galaxies, we have to know not only the distance of the host galaxy itself but the location of the host in its own galaxy cluster too.

GRBs are the most energetic transients that have been observed in the universe (Klebesadel et al. 1973; Mészáros 2006). Most of GRBs can be observed in the total range of the electromagnetic spectrum, from gamma rays to radio (Amati et al. 2013; Kulkarni et al. 1999; Mészáros et al. 2014). The duration of a GRB is characterized by the T90 parameter, which is the time taken to accumulate 90% of the total fluence registered by the detector as defined by Fishman et al. (1994). The T90 duration ranges from milliseconds to thousands of seconds, and the duration distribution includes two or more components with a main separation line around 2 s.

There are two main theories for the origin of GRBs, one with merging compact objects such as neutron stars or black holes (Berger 2014; Paczynski 1986; Usov 1992) and the other is hypernovae explosions of supermassive, low-metallicity stars (Paczyński 1998; Woosley 1993). Collision of compact objects can cause short GRBs, while hypernovae are responsible for the long ones. Long GRBs (LGRBs) can also be tracers of early star formation and, as such, may help us find large-scale structures in the universe. For example, the Hercules-Corona Borealis Great Wall by Horváth et al. (2015) and Horváth et al. (2014) and the Giant GRB Ring by Balázs et al. (2015) are the two largest known structures in the universe.

As all GRBs are at cosmological distances, and the transients decay rapidly, spectroscopic redshift measurements are very limited. We now have exact distance measurements from the afterglow of about 500 of near 10 times as more GRBs and very few of the neighboring galaxies.

Several attempts were made to give distance predictions for GRBs without knowing the exact redshift. The first predictions for mean redshifts of gamma sources were made by Horvath et al. (1996); Meszaros & Meszaros (1996, 1995); and Reichart & Mészáros (1997) from cosmological evolution and brightness distribution. Later, Bagoly et al. (2003) derived photometric redshifts of selected GRBs from peak flux ratios of different BATSE energy channels. Recently, Rácz et al. (2017) and Ukwatta et al. (2016) are derived redshifts—similar to our study—using machine-learning methods to demonstrate the underlying weak connections between the observed parameters and the distance.

To understand the physical background of LGRBs, we have to place them in their local clustering environment. Recent theories linked LGRBs to galaxies with rapid star formation or starburst (Woosley & Bloom 2006); thus, we expect that LGRBs are more frequent in midcluster galaxies where mergers and tidal interactions between gas-rich galaxies are more likely to occur.

In this paper, we show some statistical and deep-learning methods to derive redshifts from one-band optical observations for the galaxies of the hosting cluster. The available observing time for giant telescopes is heavily overbooked. The probability of obtaining approval for the submitted proposal strongly depends on the required amount of observing time. The traditional photometric z would require at least two times more observing time than the methods we used. Nevertheless, the accuracy can statistically be the same.

2 | DATA SELECTION

Photometric data were obtained by the Subaru Hyper Suprime-Cam (HSC) (Komiyama et al. 2018; Miyazaki et al. 2018) in 2017. The 1.5° field of view of the telescope with a limiting magnitude of ≈27 is ideal for searching galaxy clusters in large distances. The raw data of the r2 band (Kawanomoto et al. 2018) image was processed with the HSC v4.0.5 pipeline. The pipeline creates a database that contains the fluxes of the sources derived with several methods (from the simplest photon counting to the more complex Gaussian, Kron, and exponential model fitting algorithms) along with the area of the sources and other fitting parameters of the complex models, point spread function size, and several quality flags (Furusawa et al. 2018). In the observed image, we found 336,203 sources that were flagged as extended sources with calculated Kron and exponential flux values.

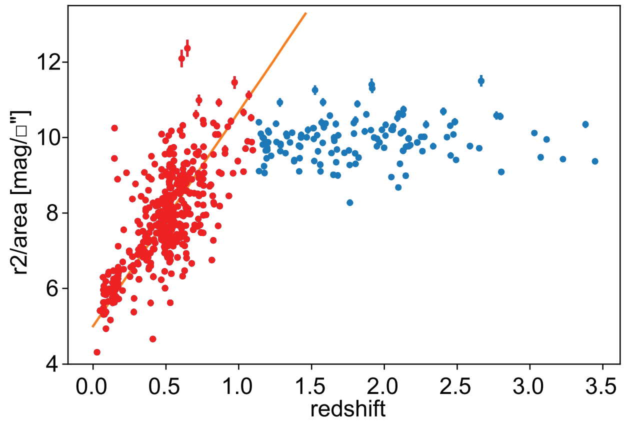

FIGURE 1 The apparent magnitude for each source divided by its area against the spectroscopic redshift. Red dots are galaxies, and blue dots are quasistellar objects

For our reference redshifts, we downloaded the Sloan Digital Sky Survey's (SDSS) (Aguado et al. 2019; Beck et al. 2016) Data Release 15 database for the area of our image. The area contains 1130 SDSS objects flagged as galaxies or quasistellar objects (QSO) with spectroscopic redshift.

Due to some large galaxies, overexposed stars, and a few defective charged coupled devices, we were only able to cross-match 814 SDSS extragalactic objects to our HSC sources. We used these 814 SDSS objects with spectroscopic redshift for our calculations.

3 | COSMOLOGICAL APPROACH

We calculated surface brightness from HSC data using the exponential model-fitting fluxes and the area derived from the fitting parameters.

Figure 1 shows that plotting surface brightness against the spectroscopic redshift demonstrated that, at higher redshifts ($z > 1$), the derived quotient is approximately constant and therefore cannot be used for redshift determination. The scattering of the data gives very large errors on prediction on lower redshifts too.

4 | PSEUDOINVERSE MATRIX

Treating the observed and derived parameters as a matrix (A) and the spectroscopic redshifts as a vector (b), one can calculate the relation between the several parameters and the redshift.

$$Ax = b \rightarrow x = A^{-1}b$$

where x is the vector of the weights of all parameters, and A^{-1} is the inverse matrix of A.

In linear algebra, a pseudoinverse (A^{+}) of a matrix A is a generalization of the inverse matrix (Ben-Israel 2003).

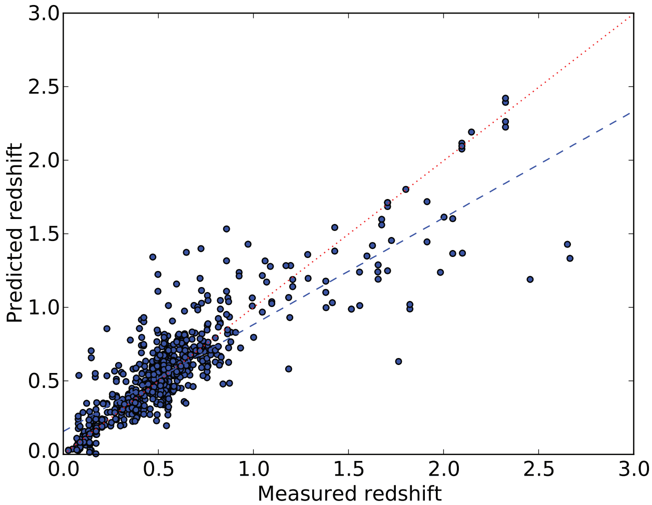

FIGURE 2 Redshifts predicted using the Moore-Penrose Pseudoinverse method versus the measured. The dashed line is the fitted

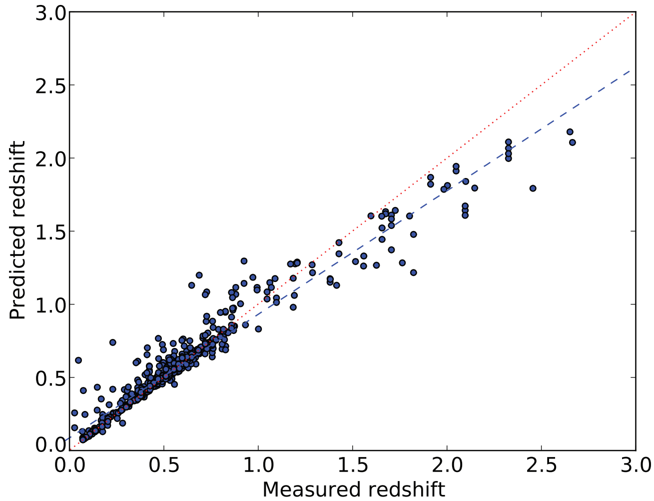

FIGURE 3 The redshifts predicted with random forests method versus the measured. The dashed line is the fitted

Because not all matrices have inverses but all matrices have pseudoinverses, in generic usage, the pseudoinverse is the best approximation for solving the system of linear equations that lacks a unique solution.

To find the pseudoinverse, we used the formula

$$A^{+} = V D^{+} U^{T}$$

where U, D, and V the left singular vectors, the singular values, and the right singular vectors of A, respectively. A^{+} is the pseudoinverse of A and D^{+} the pseudoinverse of D. D is a diagonal matrix, and thus, D^{+} can be calculated by taking the reciprocal of the nonzero values of D. More details on the pseudoinverse fitting can be found in Szécsi et al. (2013).

In Figure 2, we plotted the redshift predicted using the Moore-Penrose Pseudoinverse method versus the measured. The error of the estimation is still very high.

5 | RANDOM FORESTS

Random forests are an ensemble of learning methods for regression and classification. Several studies have used deep-learning methods to derive photometric redshifts for extragalactic sources (Beck et al. 2016; Carliles et al. 2010) but only for multiband observations.

For our study, we randomly divided the SDSS sample into a training and a test subsample. The training subsample consists of 90% of the data and the test 10%.

First, we tested our training sample with different number of trees to find which gave the least error. We found that 140 trees give the best approximation with an error of 17%.

With the parameters from the training sample, we derived the redshift values of the whole SDSS sample and then—as can be seen in Figure 3—compared them to the measured ones.

6 | SUMMARY

- In this paper, we examined several methods to predict redshifts of galaxies and QSOs from a single-band, high-angular resolution observation.
- We found that the cosmological surface brightness approach had not met our expectations as the scattering of the data means well over 50% error on low to mid redshifts and cannot predict high-redshift objects at all.
- Calculating redshift as a system of linear equations of the observed parameters lowered the error in prediction to an average of 30%. The prediction was better for low redshift objects, but with increasing redshift, the errors increase much faster.
- Using random forests methods, we further lowered the error to 17%. This value seems still too high, but it is in the same order of magnitude as the traditional multiband photo-z estimations.
- We showed that deep-learning methods can be used to derive redshifts from one-band optical observations for the galaxies of the hosting cluster.

ACKNOWLEDGMENTS

The authors thank the anonymous referees for their valuable comments that helped to improve our manuscript. This research made use of the SDSS DR15 database funded by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. This study was based on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University. This research was partly supported by the Tisza István Program, Hungary; the TKP Program, Hungary; and the OTKA grant, K131653.

AUTHOR CONTRIBUTIONS

All authors participated in the acquisition, analysis, and interpretation of data, as well as the preparation of the manuscript.

REFERENCES

Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, *ApJS*, *240*, 23.
Amati, L., Atteia, J.-L., Balazs, L. et al. 2013, *White paper, arXiv:1306.5259*.
Bagoly, Z., Csabai, I., Mészáros, A., Mészáros, P., Horváth, I., Balázs, L. G., & Vavrek, R. 2003, *A&A*, *398*, 919.
Balázs, L. G., Bagoly, Z., Hakkila, J. E., Horváth, I., Kóbori, J., Rácz, I. I., & Tóth, L. V. 2015, *MNRAS*, *452*, 2236.
Beck, R., Dobos, L., Budavári, T., Szalay, A. S., & Csabai, I. 2016, *MNRAS*, *460*, 1371.
Ben-Israel, A. 2003, *Generalized Inverses: Theory and Applications*, Springer-Verlag, New York, NY. https://books.google.hu/books?id=M8IOxQEACAAJ.
Berger, E. 2014, *ARA&A*, *52*, 43.
Carliles, S., Budavári, T., Heinis, S., Priebe, C., & Szalay, A. S. 2010, *ApJ*, *712*, 511.
Fishman, G. J., Meegan, C. A., Wilson, R. B., et al. 1994, *ApJS*, *92*, 229.
Furusawa, H., Koike, M., Takata, T., et al. 2018, *PASJ*, *70*, S3.
Horváth, I., Bagoly, Z., Hakkila, J., & Tóth, L. V. 2015, *A&A*, *584*, A48.
Horváth, I., Hakkila, J., & Bagoly, Z. 2014, *A&A*, *561*, L12.
Horvath, I., Meszaros, P., & Meszaros, A. 1996, *ApJ*, *470*, 56.
Kawanomoto, S., Uraguchi, F., Komiyama, Y., et al. 2018, *PASJ*, *70*, 66.
Klebesadel, R. W., Strong, I. B., & Olson, R. A. 1973, *ApJ*, *182*, L85.
Komiyama, Y., Obuchi, Y., Nakaya, H., et al. 2018, *PASJ*, *70*, S2.
Kulkarni, S. R., Djorgovski, S. G., Odewahn, S. C., et al. 1999, *Nature*, *398*, 389.
Meszaros, A., & Meszaros, P. 1996, *ApJ*, *466*, 29.
Mészáros, P. 2006, *Rep. Prog. Phys.*, *69*, 2259.
Mészáros, P., Asano, K., & Veres, P. 2014, *JPCS*, *485*, 012001.
Meszaros, P., & Meszaros, A. 1995, *ApJ*, *449*, 9.
Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, *PASJ*, *70*, S1.
Paczynski, B. 1986, *ApJ*, *308*, L43.
Paczyński, B. 1998, *ApJ*, *494*, L45.
Rácz, I. I., Balazs, L. G., Tóth, V. L., Horvath, I., & Bagoly, Z. 2017, Proceedings of the 7th International Fermi Symposium, 312, id. 78.
Reichart, D. E., & Mészáros, P. 1997, *ApJ*, *483*, 597.
Szécsi, D., Bagoly, Z., Kóbori, J., Horváth, I., & Balázs, L. G. 2013, *A&A*, *557*, A8.
Ukwatta, T. N., Woźniak, P. R., & Gehrels, N. 2016, *MNRAS*, *458*, 3821.
Usov, V. V. 1992, *Nature*, *357*, 472.
Woosley, S., & Bloom, J. 2006, *Annu. Rev. Astron. Astrophys*, *44*, 507.
Woosley, S. E. 1993, *ApJ*, *405*, 273.

AUTHOR BIOGRAPHY

Sandor Pinter received a B.Sc. degree in physics in 2013 (project: "Protostars in the AKARI infrared database") and an M.Sc. degree in astronomy in 2015 (based on the thesis "New methods in processing Herschel IR observations"), both under the supervision of L.V.T., Eötvös Loránd University (ELU), Budapest, Hungary. Currently, he is pursuing a Ph.D. at the School of Physics (Department of Particle Physics and Astronomy), ELU, working on "Galactic and extragalactic star formation." His research interests include statistical analysis of galactic and extragalactic star formation and separation of galaxies with different star-forming rates at various redshifts with multiwavelength observations and spectral energy distribution fitting.

How to cite this article: Pinter S, Balázs LG, Bagoly Z, Horvath I, Rácz II, Tóth VL. Methods for identifying high-redshift galaxy cluster candidates. *Astron. Nachr*. 2019;340:618–621. https://doi.org/10.1002/asna.201913665